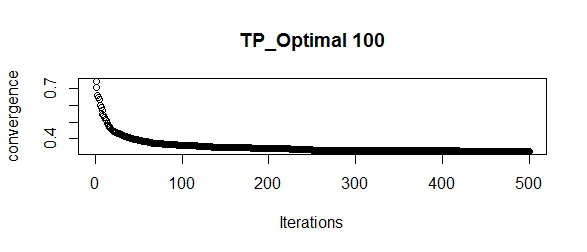
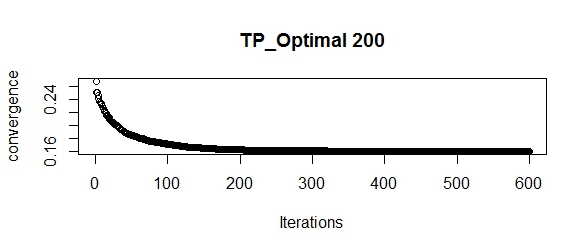
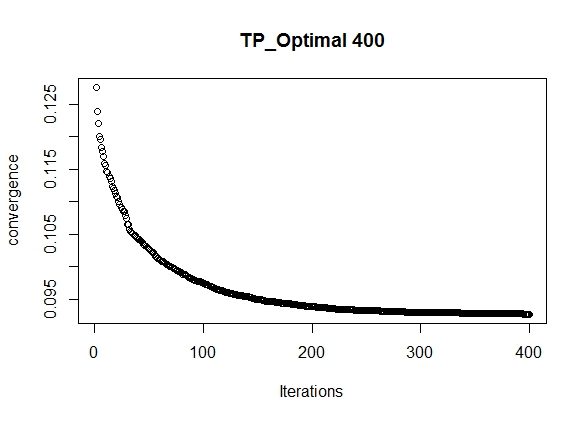
# Supplementary Figures



**Figure S1**: STPGA model convergence for optimized training population of 100 clones



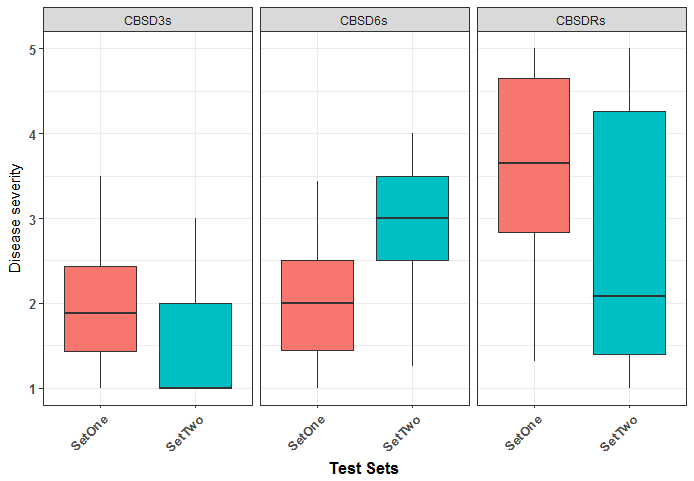
**Figure S2:** STPGA model convergence for optimized training population of 200 clones



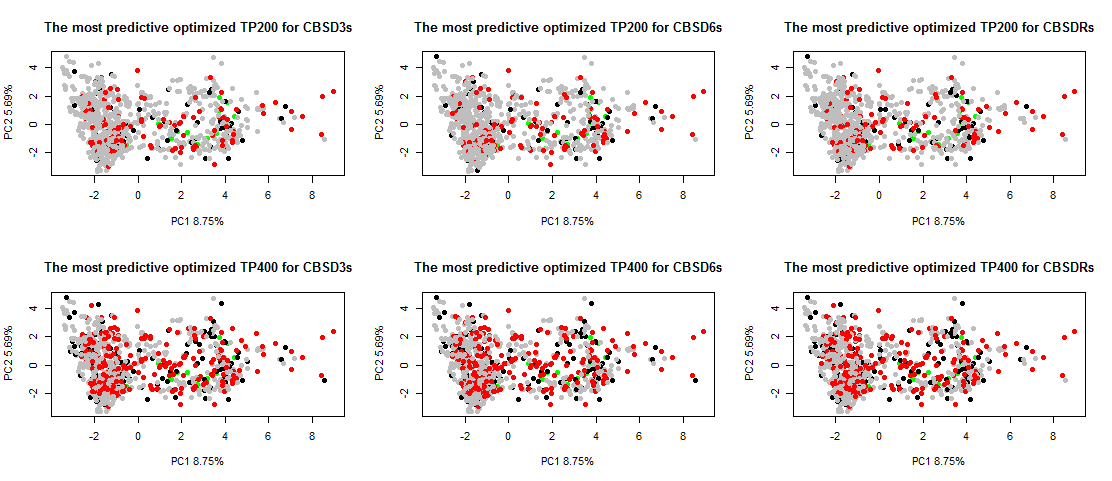
**Figure S3:** STPGA model convergence for optimized training population of 400 clones

# 

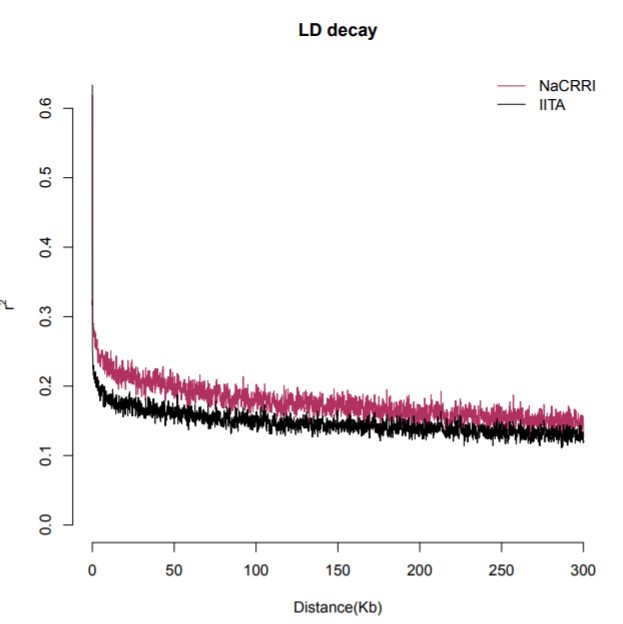
# **Figure S4**: Boxplot showing the phenotypic distribution for two training sets **(TP1 and TP2)** for the three disease traits



**Figure S5**: Boxplot showing the phenotypic distribution for the two sets of W. African clones for the three disease traits.



**Figure S6**: Plot of PC1 against PC2 for the most predictive optimized training size of 200 and 400forTP1 **(Black)** andTP2 **(Red)** as well as the unselected TP1+TP2 **(Grey)** and the IITA test set **(Green)** for the CBSD3s, CBSD6s and CBSDRs.



**Figure S7:** Linkage disequilibrium (LD) decay rate measured as the r2 values of pair-wise relationship among the markers along the chromosomes.

# Supplementary Tables

**Table S1: Prediction accuracies for optimized training population size of 100 for combined TPs (TP1 and TP2)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Training**  **Size of 100** | **G-BLUP** |  |  | **Bayes-A** |  |  | **Bayes -B** |  |  | **Bayesian Lasso** |  |  |
|  | **CBSD3s**  **Pre.acc** | **CBSD6s**  **Pre.acc** | **CBSDRs**  **Pre.acc** | **CBSD3s**  **Pre.acc** | **CBSD6s**  **Pre.acc** | **CBSDRs**  **Pre.acc** | **CBSD3s**  **Pre.acc** | **CBSD6s**  **Pre.acc** | **CBSDRs**  **Pre.acc** | **CBSD3s**  **Pre.acc** | **CBSD6s**  **Pre.acc** | **CBSDRs**  **Pre.acc** |
| TP\_1 | 0.21 | 0.21 | -0.09 | 0.22 | 0.23 | 0.03 | 0.27 | 0.24 | -0.13 | 0.31 | 0.21 | -0.15 |
| TP\_2 | 0.36 | 0.22 | -0.04 | 0.42 | 0.20 | -0.30 | 0.43 | 0.18 | -0.04 | 0.36 | 0.15 | 0.15 |
| TP\_3 | 0.09 | 0.12 | 0.00 | 0.09 | 0.11 | -0.28 | 0.07 | 0.15 | 0.02 | 0.18 | 0.15 | -0.06 |
| TP\_4 | 0.15 | 0.16 | -0.16 | 0.14 | 0.13 | -0.43 | 0.20 | 0.17 | 0.03 | 0.14 | 0.14 | -0.01 |
| TP\_5 | 0.21 | 0.22 | -0.07 | 0.23 | 0.19 | -0.11 | 0.24 | 0.21 | -0.05 | 0.29 | 0.16 | 0.04 |
| TP\_6 | 0.22 | 0.23 | -0.05 | 0.19 | 0.20 | -0.22 | 0.22 | 0.16 | 0.02 | 0.27 | 0.16 | -0.07 |
| TP\_7 | 0.44 | 0.29 | -0.19 | 0.38 | 0.25 | -0.18 | 0.43 | 0.35 | 0.09 | 0.41 | 0.27 | -0.07 |
| TP\_8 | 0.43 | 0.16 | -0.29 | 0.38 | 0.15 | -0.06 | 0.41 | 0.16 | -0.20 | 0.45 | 0.13 | -0.22 |
| TP\_9 | 0.27 | 0.28 | -0.17 | 0.31 | 0.33 | -0.31 | 0.31 | 0.32 | 0.06 | 0.39 | 0.23 | -0.06 |
| TP\_10 | 0.11 | 0.19 | -0.16 | 0.13 | 0.20 | -0.24 | 0.18 | 0.20 | -0.08 | 0.17 | 0.14 | -0.19 |
| TP\_11 | 0.37 | 0.31 | 0.10 | 0.38 | 0.24 | -0.24 | 0.49 | 0.32 | 0.08 | 0.39 | 0.19 | -0.05 |
| TP\_12 | 0.27 | 0.19 | 0.03 | 0.32 | 0.15 | -0.17 | 0.32 | 0.18 | -0.01 | 0.31 | 0.17 | -0.02 |
| TP\_13 | 0.15 | 0.28 | -0.07 | 0.23 | 0.29 | -0.32 | 0.23 | 0.27 | -0.01 | 0.28 | 0.18 | -0.08 |
| TP\_14 | 0.35 | 0.36 | -0.11 | 0.34 | 0.38 | -0.22 | 0.33 | 0.38 | 0.09 | 0.35 | 0.35 | -0.03 |
| TP\_15 | 0.08 | 0.27 | -0.07 | 0.05 | 0.28 | -0.24 | 0.15 | 0.27 | -0.11 | 0.24 | 0.22 | -0.19 |
| TP\_16 | 0.23 | 0.27 | -0.18 | 0.24 | 0.26 | -0.07 | 0.27 | 0.25 | 0.06 | 0.40 | 0.19 | -0.03 |
| TP\_17 | 0.33 | 0.20 | -0.01 | 0.33 | 0.18 | -0.21 | 0.31 | 0.18 | -0.18 | 0.40 | 0.17 | -0.14 |
| TP\_18 | 0.45 | 0.30 | -0.08 | 0.37 | 0.29 | -0.18 | 0.43 | 0.22 | -0.13 | 0.50 | 0.23 | -0.18 |
| TP\_19 | 0.23 | 0.24 | -0.18 | 0.10 | 0.27 | -0.20 | 0.26 | 0.26 | -0.05 | 0.28 | 0.19 | -0.05 |
| TP\_20 | 0.37 | 0.14 | -0.29 | 0.31 | 0.14 | 0.16 | 0.39 | 0.09 | -0.08 | 0.43 | 0.11 | 0.00 |
| **Mean Pre.acc** | **0.27** | **0.23** | **-0.10** | **0.26** | **0.22** | **-0.19** | **0.30** | **0.23** | **-0.03** | **0.33** | **0.19** | **-0.07** |

**Table S2: Prediction accuracies for optimized training population size of 200 for combined TPs (TP1 and TP2)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | |  | | |  | | |  | | |
|  | | | **G-BLUP** | |  | | |  | | | **Bayes-A** | | |  | |  | | **Bayes-B** | |  | |  | | **Bayesian Lasso** | | | |  | |  | |
| **Train. Pop size 200** | | | **M3CBSDs** | | **M6CBSDs** | | | **CBSDRs** | | | **M3CBSDs** | | | **M6CBSDs** | | **CBSDRs** | | **M3CBSDs** | | **M6CBSDs** | | **CBSDRs** | | **M3CBSDs** | | | | **M6CBSDs** | | **CBSDRs** | |
| Pre.acc (r) | | Pre.acc (r) | | | Pre.acc (r) | | | Pre.acc(r) | | | Pre.acc(r) | | Pre.acc (r) | | Pre.acc (r) | | Pre.acc (r) | | Pre.acc (r) | | Pre.acc (r) | | | | Pre.acc (r) | | Pre.acc (r) | |
| TP\_1 | | | 0.18 | | 0.33 | | | -0.02 | | | 0.20 | | | 0.31 | | -0.29 | | 0.37 | | 0.32 | | 0.18 | | 0.31 | | | | 0.28 | | 0.14 | |
| TP\_2 | | | 0.27 | | 0.38 | | | 0.06 | | | 0.33 | | | 0.36 | | -0.28 | | 0.30 | | 0.43 | | 0.13 | | 0.36 | | | | 0.34 | | 0.13 | |
| TP\_3 | | | 0.09 | | 0.34 | | | -0.03 | | | 0.10 | | | 0.34 | | -0.20 | | 0.09 | | 0.39 | | 0.10 | | 0.21 | | | | 0.20 | | 0.12 | |
| TP\_4 | | | 0.39 | | 0.29 | | | -0.05 | | | 0.40 | | | 0.32 | | -0.29 | | 0.38 | | 0.29 | | 0.14 | | 0.42 | | | | 0.27 | | 0.12 | |
| TP\_5 | | | 0.40 | | 0.25 | | | -0.04 | | | 0.24 | | | 0.27 | | -0.37 | | 0.42 | | 0.18 | | -0.10 | | 0.41 | | | | 0.21 | | -0.09 | |
| TP\_6 | | | 0.36 | | 0.27 | | | -0.08 | | | 0.36 | | | 0.24 | | -0.08 | | 0.43 | | 0.30 | | -0.03 | | 0.38 | | | | 0.18 | | -0.05 | |
| TP\_7 | | | 0.28 | | 0.29 | | | -0.03 | | | 0.24 | | | 0.25 | | -0.43 | | 0.30 | | 0.25 | | 0.15 | | 0.36 | | | | 0.21 | | 0.18 | |
| TP\_8 | | | 0.29 | | 0.20 | | | 0.01 | | | 0.29 | | | 0.16 | | -0.47 | | 0.42 | | 0.22 | | -0.03 | | 0.39 | | | | 0.22 | | 0.01 | |
| TP\_9 | | | 0.26 | | 0.36 | | | -0.12 | | | 0.29 | | | 0.36 | | -0.22 | | 0.20 | | 0.31 | | 0.13 | | 0.35 | | | | 0.29 | | 0.11 | |
| TP\_10 | | | 0.38 | | 0.37 | | | 0.07 | | | 0.42 | | | 0.35 | | -0.16 | | 0.33 | | 0.34 | | 0.00 | | 0.43 | | | | 0.22 | | -0.03 | |
| TP\_11 | | | 0.26 | | 0.29 | | | 0.11 | | | 0.21 | | | 0.27 | | -0.35 | | 0.27 | | 0.28 | | 0.12 | | 0.30 | | | | 0.23 | | 0.09 | |
| TP\_12 | | | 0.20 | | 0.31 | | | -0.05 | | | 0.27 | | | 0.31 | | -0.46 | | 0.22 | | 0.30 | | 0.22 | | 0.24 | | | | 0.22 | | 0.23 | |
| TP\_13 | | | 0.15 | | 0.14 | | | 0.00 | | | 0.23 | | | 0.15 | | -0.35 | | 0.12 | | 0.13 | | -0.12 | | 0.29 | | | | 0.13 | | -0.07 | |
| TP\_14 | | | 0.22 | | 0.22 | | | -0.02 | | | 0.23 | | | 0.21 | | -0.22 | | 0.28 | | 0.24 | | 0.10 | | 0.23 | | | | 0.21 | | 0.12 | |
| TP\_15 | | | 0.28 | | 0.31 | | | -0.09 | | | 0.23 | | | 0.23 | | -0.39 | | 0.29 | | 0.17 | | 0.10 | | 0.31 | | | | 0.23 | | 0.12 | |
| TP\_16 | | | 0.25 | | 0.17 | | | -0.09 | | | 0.27 | | | 0.09 | | -0.11 | | 0.11 | | 0.12 | | -0.09 | | 0.29 | | | | 0.13 | | -0.07 | |
| TP\_17 | | | 0.11 | | 0.19 | | | -0.12 | | | 0.07 | | | 0.18 | | -0.25 | | 0.08 | | 0.14 | | -0.02 | | 0.27 | | | | 0.15 | | -0.07 | |
| TP\_18 | | | 0.25 | | 0.23 | | | -0.09 | | | 0.27 | | | 0.19 | | -0.30 | | 0.19 | | 0.27 | | 0.13 | | 0.32 | | | | 0.19 | | 0.08 | |
| TP\_19 | | | 0.27 | | 0.21 | | | -0.09 | | | 0.30 | | | 0.19 | | -0.31 | | 0.33 | | 0.20 | | 0.10 | | 0.36 | | | | 0.21 | | 0.05 | |
| TP\_20 | | | 0.38 | | 0.27 | | | 0.02 | | | 0.25 | | | 0.33 | | -0.26 | | 0.37 | | 0.35 | | 0.13 | | 0.31 | | | | 0.26 | | 0.15 | |
| **Mean Pre.acc** | | | **0.27** | | **0.28** | | | **-0.03** | | | **0.26** | | | **0.26** | | **-0.29** | | **0.27** | | **0.26** | | **0.07** | | **0.34** | | | | **0.22** | | **0.06** | |
| **Table S3: Prediction accuracies for optimized training population size of 400 for combined TPs** | | | | | | | | | | | | | | | | | | | | | | | | | |  | |  | |  | |  | | |
| Training pop 400 | | **G-BLUP** | | | | |  | | |  | | | **Bayes-A** | | |  | |  | | **Bayes-B** | |  | |  | | | **Bayesian Lasso** | | |  | |  | |
|  | | **CBSD3s**  **Pred.acc** | | | | | **CBSD6s**  **Pred.acc** | | | **CBSDRs**  **Pred.acc** | | | **CBSD3s**  **Pred.acc** | | | **CBSD6s**  **Pred.acc** | | **CBSDRs**  **Pred.acc** | | **CBSD3s**  **Pred.acc** | | **CBSD6s**  **Pred.acc** | | **CBSDRs**  **Pred.acc** | | | **CBSD3s**  **Pred.acc** | | | **CBSD6s**  **Pred.acc** | | **CBSDRs**  **Pred.acc** | |
| TP\_1 | | 0.40 | | | | | 0.18 | | | 0.01 | | | 0.40 | | | 0.18 | | -0.30 | | 0.42 | | 0.14 | | -0.12 | | | 0.46 | | | 0.13 | | -0.11 | |
| TP\_2 | | 0.28 | | | | | 0.19 | | | 0.06 | | | 0.26 | | | 0.21 | | -0.11 | | 0.32 | | 0.16 | | -0.05 | | | 0.34 | | | 0.12 | | -0.07 | |
| TP\_3 | | 0.32 | | | | | 0.16 | | | 0.01 | | | 0.32 | | | 0.16 | | -0.10 | | 0.33 | | 0.12 | | -0.11 | | | 0.38 | | | 0.12 | | -0.10 | |
| TP\_4 | | 0.34 | | | | | 0.21 | | | -0.05 | | | 0.33 | | | 0.18 | | -0.10 | | 0.31 | | 0.24 | | -0.08 | | | 0.33 | | | 0.12 | | -0.16 | |
| TP\_5 | | 0.32 | | | | | 0.19 | | | -0.03 | | | 0.30 | | | 0.18 | | -0.20 | | 0.37 | | 0.19 | | -0.12 | | | 0.39 | | | 0.14 | | -0.12 | |
| TP\_6 | | 0.19 | | | | | 0.21 | | | -0.02 | | | 0.10 | | | 0.22 | | -0.23 | | 0.23 | | 0.17 | | -0.04 | | | 0.25 | | | 0.13 | | -0.05 | |
| TP\_7 | | 0.34 | | | | | 0.18 | | | 0.04 | | | 0.38 | | | 0.19 | | -0.25 | | 0.35 | | 0.22 | | -0.09 | | | 0.37 | | | 0.20 | | -0.04 | |
| TP\_8 | | 0.26 | | | | | 0.17 | | | -0.04 | | | 0.30 | | | 0.22 | | -0.17 | | 0.26 | | 0.22 | | 0.01 | | | 0.35 | | | 0.16 | | -0.02 | |
| TP\_9 | | 0.34 | | | | | 0.19 | | | 0.00 | | | 0.36 | | | 0.15 | | -0.16 | | 0.30 | | 0.13 | | -0.08 | | | 0.40 | | | 0.13 | | -0.08 | |
| TP\_10 | | 0.35 | | | | | 0.23 | | | -0.05 | | | 0.33 | | | 0.20 | | -0.28 | | 0.34 | | 0.23 | | -0.11 | | | 0.36 | | | 0.16 | | -0.15 | |
| TP\_11 | | 0.38 | | | | | 0.17 | | | -0.04 | | | 0.37 | | | 0.16 | | -0.22 | | 0.36 | | 0.17 | | -0.13 | | | 0.38 | | | 0.14 | | -0.13 | |
| TP\_12 | | 0.26 | | | | | 0.14 | | | 0.01 | | | 0.33 | | | 0.13 | | -0.11 | | 0.22 | | 0.10 | | -0.08 | | | 0.32 | | | 0.10 | | -0.05 | |
| TP\_13 | | 0.39 | | | | | 0.23 | | | 0.00 | | | 0.43 | | | 0.20 | | -0.32 | | 0.43 | | 0.19 | | -0.12 | | | 0.43 | | | 0.13 | | -0.11 | |
| TP\_14 | | 0.24 | | | | | 0.19 | | | 0.02 | | | 0.27 | | | 0.18 | | -0.23 | | 0.27 | | 0.18 | | -0.11 | | | 0.35 | | | 0.14 | | -0.08 | |
| TP\_15 | | 0.31 | | | | | 0.21 | | | -0.06 | | | 0.30 | | | 0.23 | | -0.11 | | 0.34 | | 0.13 | | -0.13 | | | 0.32 | | | 0.13 | | -0.04 | |
| TP\_16 | | 0.31 | | | | | 0.20 | | | 0.00 | | | 0.32 | | | 0.21 | | -0.24 | | 0.34 | | 0.20 | | -0.08 | | | 0.34 | | | 0.15 | | -0.02 | |
| TP\_17 | | 0.24 | | | | | 0.20 | | | 0.01 | | | 0.26 | | | 0.17 | | -0.08 | | 0.22 | | 0.17 | | -0.01 | | | 0.31 | | | 0.18 | | -0.05 | |
| TP\_18 | | 0.37 | | | | | 0.20 | | | 0.01 | | | 0.42 | | | 0.19 | | -0.20 | | 0.41 | | 0.18 | | -0.06 | | | 0.41 | | | 0.13 | | -0.09 | |
| TP\_19 | | 0.34 | | | | | 0.17 | | | -0.08 | | | 0.26 | | | 0.17 | | -0.13 | | 0.26 | | 0.16 | | -0.15 | | | 0.37 | | | 0.12 | | -0.11 | |
| TP\_20 | | 0.25 | | | | | 0.16 | | | -0.04 | | | 0.30 | | | 0.14 | | -0.25 | | 0.26 | | 0.16 | | -0.11 | | | 0.30 | | | 0.12 | | -0.07 | |
| **Mean** | | **0.32** | | | | | **0.19** | | | **-0.01** | | | **0.32** | | | **0.18** | | **-0.19** | | **0.32** | | **0.17** | | **-0.09** | | | **0.36** | | | **0.14** | | **-0.08** | |

**Table S4: Prediction accuracies for optimized training population size of 800 for TP1 and TP2**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **G-BLUP** |  |  | **Bayes-A** |  |  | **Bayes-B** |  |  | **Bayesian Lasso** |  |  |
| **Train. Pop size 800** | **M3CBSDs** | **M6CBSDs** | **CBSDRs** | **M3CBSDs** | **M6CBSDs** | **CBSDRs** | **M3CBSDs** | **M6CBSDs** | **CBSDRs** | **M3CBSDs** | **M6CBSDs** | **CBSDRs** |
| Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc(r) | Pre.acc(r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) |
| TP\_1 | 0.30 | 0.25 | 0.05 | 0.30 | 0.21 | -0.13 | 0.27 | 0.21 | -0.09 | 0.29 | 0.19 | -0.01 |
| TP\_2 | 0.29 | 0.25 | 0.07 | 0.31 | 0.25 | -0.09 | 0.27 | 0.25 | -0.05 | 0.29 | 0.17 | 0.02 |
| TP\_3 | 0.31 | 0.26 | 0.06 | 0.34 | 0.26 | -0.10 | 0.27 | 0.20 | -0.05 | 0.33 | 0.15 | 0.00 |
| TP\_4 | 0.31 | 0.26 | 0.07 | 0.28 | 0.25 | -0.14 | 0.30 | 0.25 | 0.03 | 0.30 | 0.17 | 0.03 |
| TP\_5 | 0.30 | 0.27 | 0.07 | 0.12 | 0.21 | -0.13 | 0.28 | 0.27 | -0.04 | 0.29 | 0.13 | 0.00 |
| TP\_6 | 0.30 | 0.27 | 0.06 | 0.33 | 0.26 | -0.17 | 0.12 | 0.26 | -0.05 | 0.29 | 0.18 | -0.04 |
| TP\_7 | 0.30 | 0.27 | 0.07 | 0.35 | 0.22 | -0.18 | 0.27 | 0.25 | -0.01 | 0.27 | 0.17 | 0.00 |
| TP\_8 | 0.29 | 0.27 | 0.06 | 0.32 | 0.24 | -0.15 | 0.30 | 0.28 | -0.02 | 0.30 | 0.21 | 0.00 |
| TP\_9 | 0.30 | 0.26 | 0.07 | 0.33 | 0.27 | -0.11 | 0.21 | 0.14 | -0.05 | 0.28 | 0.16 | 0.00 |
| TP\_10 | 0.31 | 0.26 | 0.07 | 0.34 | 0.25 | -0.20 | 0.38 | 0.26 | -0.07 | 0.34 | 0.19 | -0.05 |
| TP\_11 | 0.30 | 0.28 | 0.05 | 0.32 | 0.29 | -0.13 | 0.35 | 0.33 | 0.01 | 0.32 | 0.16 | -0.01 |
| TP\_12 | 0.30 | 0.26 | 0.07 | 0.34 | 0.27 | -0.11 | 0.37 | 0.23 | -0.03 | 0.34 | 0.18 | 0.01 |
| TP\_13 | 0.30 | 0.26 | 0.07 | 0.16 | 0.25 | -0.12 | 0.34 | 0.13 | -0.05 | 0.28 | 0.17 | -0.03 |
| TP\_14 | 0.30 | 0.28 | 0.07 | 0.16 | 0.23 | -0.20 | 0.33 | 0.21 | -0.05 | 0.36 | 0.20 | -0.03 |
| TP\_15 | 0.30 | 0.26 | 0.07 | 0.30 | 0.27 | 0.12 | 0.24 | 0.24 | 0.04 | 0.30 | 0.12 | -0.01 |
| TP\_16 | 0.30 | 0.27 | 0.06 | 0.26 | 0.26 | -0.18 | 0.30 | 0.30 | -0.04 | 0.30 | 0.20 | 0.00 |
| TP\_17 | 0.30 | 0.26 | 0.06 | 0.31 | 0.25 | -0.19 | 0.32 | 0.18 | -0.10 | 0.37 | 0.16 | -0.01 |
| TP\_18 | 0.30 | 0.27 | 0.06 | 0.34 | 0.26 | -0.11 | 0.25 | 0.33 | -0.05 | 0.30 | 0.19 | -0.01 |
| TP\_19 | 0.29 | 0.25 | 0.05 | 0.29 | 0.25 | -0.07 | 0.28 | 0.15 | -0.07 | 0.30 | 0.13 | -0.02 |
| TP\_20 | 0.28 | 0.25 | 0.06 | 0.26 | 0.23 | -0.18 | 0.26 | 0.18 | -0.05 | 0.29 | 0.11 | -0.01 |
| **Mean Pre.acc** | **0.31** | **0.26** | **0.06** | **0.29** | **0.25** | **-0.13** | **0.29** | **0.23** | **-0.04** | **0.31** | **0.17** | **-0.01** |

**Table S5: Prediction accuracies for full set (922) of clones (TP1 and TP2)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | |  |  |  |  |  |  |  |
|  | **G-BLUP** |  |  | **Bayes-A** |  |  | **Bayes-B** |  |  | **Bayesian Lasso** |  |  |
| **Train. Pop size 922** | **M3CBSDs** | **M6CBSDs** | **CBSDRs** | **M3CBSDs** | **M6CBSDs** | **CBSDRs** | **M3CBSDs** | **M6CBSDs** | **CBSDRs** | **M3CBSDs** | **M6CBSDs** | **CBSDRs** |
| Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc(r) | Pre.acc(r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) | Pre.acc (r) |
| **Pre.acc** | **0.30** | **0.25** | **0.05** | **0.24** | **0.21** | **-0.11** | **0.30** | **0.26** | **-0.09** | **0.31** | **0.15** | **-0.04** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table S6: Comparing prediction accuracies for optimized and random subset of training population of size 200** | | | | | | |
|  |  | **Optimized** |  |  | **Random sub**  **Set of TP** |  |
| **Subset of TP** |
| **Train. Pop size 200** | **CBSD3s** | **CBSD6s** | **CBSDRs** | **CBSD3s** | **CBSD6s** | **CBSDRs** |
| TP\_1 | 0.18 | 0.33 | -0.42 | 0.30 | 0.01 | -0.02 |
| TP\_2 | 0.27 | 0.38 | -0.33 | 0.06 | 0.00 | 0.06 |
| TP\_3 | 0.09 | 0.34 | 0.08 | 0.40 | 0.15 | -0.03 |
| TP\_4 | 0.39 | 0.29 | 0.33 | 0.25 | 0.06 | -0.05 |
| TP\_5 | 0.40 | 0.25 | 0.16 | 0.36 | -0.04 | -0.04 |
| TP\_6 | 0.36 | 0.27 | 0.16 | 0.16 | -0.03 | -0.08 |
| TP\_7 | 0.28 | 0.29 | 0.30 | 0.50 | 0.04 | -0.03 |
| TP\_8 | 0.29 | 0.20 | 0.05 | 0.15 | -0.23 | 0.01 |
| TP\_9 | 0.26 | 0.36 | -0.17 | -0.18 | -0.06 | -0.12 |
| TP\_10 | 0.38 | 0.37 | -0.17 | 0.07 | -0.07 | 0.07 |
| TP\_11 | 0.26 | 0.29 | -0.11 | -0.07 | 0.07 | 0.11 |
| TP\_12 | 0.20 | 0.31 | -0.12 | -0.11 | 0.01 | -0.05 |
| TP\_13 | 0.15 | 0.14 | -0.16 | 0.06 | -0.17 | 0.00 |
| TP\_14 | 0.22 | 0.22 | -0.13 | 0.34 | -0.04 | -0.02 |
| TP\_15 | 0.28 | 0.31 | -0.15 | -0.13 | 0.03 | -0.09 |
| TP\_16 | 0.25 | 0.17 | 0.18 | -0.12 | 0.01 | -0.09 |
| TP\_17 | 0.11 | 0.19 | 0.02 | -0.07 | -0.19 | -0.12 |
| TP\_18 | 0.25 | 0.23 | -0.18 | -0.13 | 0.06 | -0.09 |
| TP\_19 | 0.27 | 0.21 | -0.19 | 0.23 | 0.09 | -0.09 |
| TP\_20 | 0.38 | 0.27 | 0.12 | 0.21 | 0.11 | 0.02 |
| **Mean Pre.acc** | **0.27** | **0.28** | **-0.04** | **0.11** | **-0.01** | **-0.03** |

**Table S7: Comparing prediction accuracies for optimized and random subset of training population of size 400**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  |  | **Optimized** |  |  | **Random sub** |  |
| **Subset of TP** | **Set of TP** |
| **Train. Pop size 400** | **CBSD3s** | **CBSD6s** | **CBSDRs** | **CBSD3s** | **CBSD6s** | **CBSDRs** |
| TP\_1 | 0.40 | 0.18 | 0.01 | 0.17 | -0.04 | -0.03 |
| TP\_2 | 0.28 | 0.19 | 0.06 | 0.05 | 0.05 | 0.03 |
| TP\_3 | 0.32 | 0.16 | 0.01 | 0.15 | 0.14 | -0.02 |
| TP\_4 | 0.34 | 0.21 | -0.05 | 0.02 | 0.00 | 0.04 |
| TP\_5 | 0.32 | 0.19 | -0.03 | 0.07 | 0.06 | 0.17 |
| TP\_6 | 0.19 | 0.21 | -0.02 | -0.02 | 0.16 | -0.21 |
| TP\_7 | 0.34 | 0.18 | 0.04 | 0.19 | -0.01 | 0.22 |
| TP\_8 | 0.26 | 0.17 | -0.04 | 0.29 | 0.01 | -0.01 |
| TP\_9 | 0.34 | 0.19 | 0.00 | -0.10 | 0.10 | -0.21 |
| TP\_10 | 0.35 | 0.23 | -0.05 | 0.26 | 0.12 | -0.12 |
| TP\_11 | 0.38 | 0.17 | -0.04 | 0.38 | 0.11 | 0.07 |
| TP\_12 | 0.26 | 0.14 | 0.01 | 0.08 | -0.08 | -0.03 |
| TP\_13 | 0.39 | 0.23 | 0.00 | -0.04 | -0.12 | 0.11 |
| TP\_14 | 0.24 | 0.19 | 0.02 | 0.24 | 0.20 | -0.02 |
| TP\_15 | 0.31 | 0.21 | -0.06 | 0.17 | 0.09 | 0.04 |
| TP\_16 | 0.31 | 0.20 | 0.00 | 0.02 | -0.05 | 0.08 |
| TP\_17 | 0.24 | 0.20 | 0.01 | 0.02 | 0.01 | -0.02 |
| TP\_18 | 0.37 | 0.20 | 0.01 | 0.03 | 0.07 | -0.28 |
| TP\_19 | 0.34 | 0.17 | -0.08 | 0.23 | 0.01 | 0.08 |
| TP\_20 | 0.25 | 0.16 | -0.04 | -0.14 | -0.11 | 0.01 |
| **Mean Pre.acc** | **0.32** | **0.19** | **-0.01** | **0.10** | **0.04** | **-0.01** |

**Table S8: Prediction accuracies for single and multi-kernel G-BLUP models for optimized training population size of 200 clones, where K\_1, K\_2 and K\_3 represent single kernel, two kernels, and three kernels G-BLUP models respectively. K\_1 is fitted with one genomic relationship matrix, K\_2 with two genomic relationship matrix is characterized by the first genomic relationship matrix arising from combined CBSD QTL markers on chromosomes 4 and 11, and second kernel being the rest of the genomic markers, while K\_3 with 3 genomic relationship matrices is characterized by the first two genomic relationship matrices arising from CBSD QTL markers on chromosomes 4 and 11, and third kernel being the rest of the genomic markers**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Training set 200** | **CBSD3s** | | | **CBSD6s** | | |  | **CBSDRs** | | |
| K\_1 | K\_2 | K\_3 | K\_1 | K\_2 | K\_3 |  | K\_1 | K\_2 | K\_3 |
| TP1 | 0.18 | 0.27 | 0.30 | 0.34 | 0.44 | 0.45 |  | -0.02 | -0.01 | 0.09 |
| TP2 | 0.27 | 0.29 | 0.32 | 0.39 | 0.46 | 0.42 |  | 0.06 | 0.07 | -0.02 |
| TP3 | 0.09 | 0.47 | 0.42 | 0.35 | 0.19 | 0.47 |  | -0.03 | -0.07 | -0.06 |
| TP4 | 0.38 | 0.42 | 0.43 | 0.29 | 0.45 | 0.38 |  | -0.05 | -0.06 | -0.09 |
| TP5 | 0.40 | 0.45 | 0.46 | 0.26 | 0.34 | 0.36 |  | -0.04 | -0.07 | -0.15 |
| TP6 | 0.35 | 0.31 | 0.33 | 0.29 | 0.34 | 0.41 |  | -0.08 | -0.21 | -0.02 |
| TP7 | 0.37 | 0.32 | 0.27 | 0.20 | 0.40 | 0.29 |  | -0.03 | -0.01 | -0.03 |
| TP8 | 0.29 | 0.30 | 0.35 | 0.37 | 0.30 | 0.50 |  | 0.01 | -0.02 | -0.12 |
| TP9 | 0.26 | 0.35 | 0.26 | 0.37 | 0.49 | 0.47 |  | -0.12 | 0.11 | -0.04 |
| TP10 | 0.37 | 0.30 | 0.21 | 0.30 | 0.47 | 0.43 |  | 0.07 | -0.06 | 0.13 |
| TP11 | 0.26 | 0.23 | 0.25 | 0.31 | 0.42 | 0.44 |  | 0.11 | 0.12 | -0.05 |
| TP12 | 0.20 | 0.19 | 0.22 | 0.15 | 0.27 | 0.28 |  | -0.05 | -0.02 | -0.03 |
| TP13 | 0.15 | 0.22 | 0.23 | 0.23 | 0.32 | 0.28 |  | 0.00 | 0.00 | 0.02 |
| TP14 | 0.22 | 0.31 | 0.32 | 0.31 | 0.43 | 0.41 |  | -0.02 | -0.12 | -0.11 |
| TP15 | 0.28 | 0.27 | 0.22 | 0.17 | 0.17 | 0.17 |  | -0.09 | -0.15 | -0.11 |
| TP16 | 0.24 | 0.43 | 0.39 | 0.20 | 0.14 | 0.30 |  | -0.09 | -0.13 | -0.12 |
| TP17 | 0.11 | 0.34 | 0.34 | 0.23 | 0.33 | 0.26 |  | -0.12 | -0.15 | -0.15 |
| TP18 | 0.25 | 0.29 | 0.41 | 0.21 | 0.50 | 0.30 |  | -0.09 | -0.13 | -0.12 |
| TP19 | 0.26 | 0.20 | 0.26 | 0.38 | 0.25 | 0.33 |  | -0.09 | -0.13 | -0.12 |
| TP20 | 0.27 | 0.25 | 0.30 | 0.22 | 0.31 | 0.51 |  | 0.02 | -0.01 | 0.24 |
| **Mean Pred** | **0.27** | **0.31** | **0.32** | **0.28** | **0.35** | **0.37** |  | **-0.03** | **-0.05** | **-0.04** |

**Table S9: Prediction accuracies for single and multi-kernel G-BLUP models for optimized training population of size 400, K\_1, K\_2 and K\_3 represent single kernel, two kernels, and three kernels G-BLUP models respectively. K\_1 was fitted with one genomic relationship matrix, K\_2 with two genomic relationship matrix was characterized by the first genomic relationship matrix arising from combined CBSD QTL markers on chromosomes 4 and 11, and second kernel being the rest of the genomic markers, while K\_3 with 3 genomic relationship matrices was characterized by the first two genomic relationship matrices arising from CBSD QTL markers on chromosomes 4 and 11, and third kernel being the rest of the genomic markers**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Training set 400 | **CBSD3s** | | | **CBSD6s** | | | **CBSDRs** | | |
| K\_1 | K\_2 | K\_3 | K\_1 | K\_2 | K\_3 | K\_1 | K\_2 | K\_3 |
| TP1 | 0.40 | 0.42 | 0.41 | 0.17 | 0.24 | 0.26 | 0.00 | 0.01 | 0.02 |
| TP2 | 0.28 | 0.24 | 0.23 | 0.19 | 0.34 | 0.26 | 0.06 | -0.12 | -0.10 |
| TP3 | 0.32 | 0.39 | 0.34 | 0.16 | 0.23 | 0.25 | 0.01 | 0.00 | 0.01 |
| TP4 | 0.34 | 0.31 | 0.34 | 0.21 | 0.27 | 0.29 | -0.01 | -0.09 | -0.04 |
| TP5 | 0.32 | 0.27 | 0.29 | 0.19 | 0.25 | 0.27 | -0.05 | -0.05 | -0.03 |
| TP6 | 0.19 | 0.43 | 0.43 | 0.21 | 0.30 | 0.30 | -0.03 | -0.04 | -0.03 |
| TP7 | 0.34 | 0.31 | 0.37 | 0.17 | 0.23 | 0.26 | -0.02 | 0.03 | 0.04 |
| TP8 | 0.26 | 0.24 | 0.28 | 0.17 | 0.27 | 0.28 | 0.04 | -0.07 | -0.08 |
| TP9 | 0.34 | 0.29 | 0.36 | 0.19 | -0.24 | 0.29 | -0.04 | -0.01 | 0.00 |
| TP10 | 0.35 | 0.32 | 0.36 | 0.23 | 0.28 | 0.30 | 0.00 | -0.05 | -0.03 |
| TP11 | 0.38 | 0.28 | 0.36 | 0.24 | 0.23 | 0.27 | -0.05 | -0.07 | -0.04 |
| TP12 | 0.26 | 0.27 | 0.28 | 0.14 | 0.20 | 0.23 | -0.04 | 0.00 | 0.01 |
| TP13 | 0.39 | 0.34 | 0.42 | 0.22 | 0.29 | 0.31 | 0.01 | 0.00 | 0.01 |
| TP14 | 0.24 | 0.34 | 0.24 | 0.19 | 0.25 | 0.31 | 0.00 | 0.01 | 0.01 |
| TP15 | 0.31 | 0.32 | 0.33 | 0.21 | 0.28 | 0.29 | 0.02 | -0.01 | -0.05 |
| TP16 | 0.31 | 0.29 | 0.32 | 0.20 | 0.26 | 0.28 | -0.06 | 0.00 | 0.01 |
| TP17 | 0.24 | 0.29 | 0.28 | 0.20 | 0.23 | 0.29 | 0.00 | 0.02 | 0.04 |
| TP18 | 0.37 | 0.39 | 0.40 | 0.20 | 0.24 | 0.26 | 0.01 | 0.00 | 0.01 |
| TP19 | 0.34 | 0.29 | 0.30 | 0.17 | 0.20 | 0.27 | 0.01 | -0.08 | -0.06 |
| TP20 | 0.25 | 0.34 | 0.49 | 0.16 | 0.25 | 0.23 | -0.08 | -0.01 | -0.01 |
| Mean Pred. | **0.32** | **0.32** | **0.34** | **0.19** | **0.23** | **0.27** | **-0.01** | **-0.03** | **-0.02** |

**Table S10: Five-fold cross validation, replicated 10 times for IITA clones G-BLUP model**

|  |  |  |  |
| --- | --- | --- | --- |
| **Replications** | **CBSD3s** | **CBSD6s** | **CBSDRs** |
| 1 | 0.39 | 0.23 | 0.08 |
| 2 | 0.45 | 0.24 | 0.04 |
| 3 | 0.41 | 0.42 | 0.22 |
| 4 | 0.37 | 0.21 | 0.20 |
| 5 | 0.46 | 0.34 | -0.03 |
| 6 | 0.37 | -0.29 | 0.13 |
| 7 | 0.48 | 0.27 | 0.07 |
| 8 | 0.45 | 0.27 | -0.20 |
| 9 | 0.36 | 0.22 | 0.13 |
| 10 | 0.33 | 0.14 | 0.18 |
| **Mean Pred.** | **0.40** | **0.21** | **0.08** |

**Table S11: Prediction accuracies of CBSD-traits for single and multi-kernel G-BLUP models under high density, whole genome sequence imputed markers (WGS) and low density genotyping-by sequencing markers (GBS) markers for optimized training population size of 200 clones.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Traits | **Single Kernel model (K1)** | | **Multi-Kernel (K2)** | | **Multi-Kernel (K3)** | |
| GBS markers | WGS Markers | GBS markers | WGS Markers | GBS markers | WGS Markers |
| CBSD3s | 0.27 | 0.35 | 0.31 | 0.26 | 0.32 | 0.32 |
| CBSD6s | 0.28 | 0.15 | 0.35 | 0.24 | 0.37 | 0.22 |
| CBSDRs | -0.03 | 0.18 | -0.05 | 0.19 | -0.04 | 0.14 |

**Table S12: Prediction accuracies for CBSD related traits for single and multi-kernel G-BLUP models under high density, whole genome sequence imputed markers (WGS) and low density genotyping-by sequencing markers (GBS) markers for optimized training population size of 400 clones**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Traits | **Single Kernel model (K1)** | | **Multi-Kernel (K2)** | | **Multi-Kernel (K3)** | |
| GBS markers | WGS Markers | GBS markers | WGS Markers | GBS markers | WGS Markers |
| CBSD3s | 0.32 | 0.39 | 0.34 | 0.41 | 0.44 | 0.32 |
| CBSD6s | 0.19 | 0.15 | 0.23 | 0.16 | 0.20 | 0.27 |
| CBSDRs | -0.01 | 0.16 | -0.03 | 0.15 | 0.17 | -0.02 |

**Table S13: Variance component and heritability estimates of TP1 and TP2 for CBSD-traits**.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Datasets** |  | | **TP1** |  |  |  |  | **TP2** |  |
| **Sources Variations** | | **CBSD3s** | **CBSD6s** | **CBSDRs** |  | **Sources Variations** | **CBSD3s** | **CBSD6s** | **CBSDRs** |
| Rep/Loc | 0.127 | | 0.013 | 0.017 |  | Block/Loc | 0.017 | 0.059 | 0.009 |
| Clones | 0.132 | | 0.228 | 0.453 |  | Clones | 0.173 | 0.213 | 0.318 |
| Clones x Loc | 0.025 | | 0.056 | 0.42 |  | Clones x Loc | 0.008 | 0.119 | 0.096 |
| Residual | 0.34 | | 0.446 | 0.64 |  | Residual | 0.385 | 0.529 | 0.471 |
| **H2** | **0.28** | | **0.34** | **0.42** |  | **H2** | **0.31** | **0.29** | **0.40** |